Reviewer 1

Many thanks for your constructive comments on our manuscript. Please find our responses below, with your original comments in regular text and our responses underneath in green:

'The N14CP model isn't described in much detail, but is reported to have been validated against an extensive range of sites. It appears to be a largely empirical, first-order mathematical approach with climate as the primary plant growth driver, limited by N and/or P availability to meet stoichiometric needs. The rate equations for P uptake and loss are not defined clearly, but the new "cleaving parameter" appears to be a rate coefficient for a first-order model.'

We appreciate your comment. Rather than repeat previously published detail we refer to other papers that include more in-depth description as is often the convention with process based modelling. Instead we have attempted to summarise the most pertinent aspects and your interpretation of the N14CP model and the cleaving parameter from the detail provided is accurate. We have perhaps missed some relevant detail in doing so and as you suggest, we could more clearly define the rate equations for P uptake and loss – thank you for this suggestion, we will happily provide this detail from the original model development paper here to aid the reader.

This detail will be added in the '2.2.1. N14CP model summary' section where we discuss plant sources of nutrients

'The authors finally mention root surface phosphatase enzyme activities on line 539. This should have been done much sooner as a justification for the model formulation and perhaps help inform model formulation. In all, it's difficult to understand how the model works from the information provided.'

We agree that explicit and earlier reference to the relevance of the $P_{CleaveMax}$ parameter to root surface phosphatase activity would be useful.

This shall be added and clarified in '2.3.3. Model parameters for the acidic and calcareous grasslands' section.

'The principle focus of simulations seemed to be to derive optimal parameter values that provided best fit to a set of field observations for aboveground biomass (AGB), and soil organic C, N and P pools. The comparisons were difficult to interpret given the scales of simulated vs observed data (Fig. 2).'

While deriving optimal parameter values was an important component of the simulations, the principle focus of the simulations was to explore variation in P acquisition parameters and how these may help account for differing responses to N or P addition. We attempt to find parameter values for these uncertain and under-studied processes by using the observational data from this long-term manipulation experiment.

We appreciate that this focus may be unclear and so will add some clarification in the aims section of the introduction.

With respect to the scales – in these 1:1s we believe retaining the 0 at the lower limit is important as it provides a scale context. We agree however, perhaps the upper limits could be reduced somewhat to narrow the scale slightly and enhance readability.

'Although the overall relationship between simulated and observed AGB was reasonable, these relationships for soil pools were much weaker and would be better understood if scales were selected to spread the observations.'

As explained above, we decided to plot raw (untransformed) observational data against simulated data with x and y axes beginning at 0 to present data as transparently as possible. We previously plotted all data across all variables and both grasslands on the same natural log-scaled axes, which whilst helping the visual spread of data, made the relationship between simulated and observed data appear much more close than in reality, and hence we decided not to use such a format.

We will narrow the upper scales for densely clustered variables (such as SOC and SON) in figure 2 to aid readability as suggested above.

'From these data, there was no apparent relationship between observed and simulated values of C and P'

We acknowledge that in its current state, figure 2 may not be adequately highlighting relationships between observed and simulated values of C and P. We hope that with amendments to the scales and other adjustments that these relationships may be clearer. Where amendments to figure 2 don't help clarify the relationships, we could add more explanation in text.

The relationships that we had hoped to highlight are as follows:

The differences between SOC in the acid plots are similar for simulated and observed, but the magnitude of the estimated pool is higher overall. The model captures the pattern of increasing SOC with N addition fairly well, though for reasons we identify in the manuscript, does not pick up on the increase in SOC with P addition.

For the calcareous plots the magnitude is better captured for SOC and the higher SOC in the P addition plot is picked out by the model. However, unlike with the acidic SOC, the relationship between calcareous SOC and N treatment is less clearly identified, though it is worth noting that there is less difference in the observed SOCs for this site. Taken together, these produce a (misleadingly) small r^2 value of 0.01, which would be considerably larger if grasslands where plotted separately.

For P – the magnitudes of the calcareous sites are well captured and there is little difference between the two plots observed. We agree that there is discrepancy between the observed and simulated for the P in the acid site. Why we believe such a discrepancy exists is discussed in the discussion section (lines 508 to 511), and we formally quantify the relationships using r^2 and errors in the results section.

‘Given the weak validation test results, the lengthy discussions of many simulation patterns and details seem overemphasized. Many of the results are highly speculative and for reasons that aren’t clear. For example, the authors discuss many potential interactions between N and P limitations, but how are these explanations based on mechanisms included in the model?’

We hope we have somewhat addressed the overemphasis of this section of the discussion in our previous comment. We do agree that this section is perhaps too lengthy and as you identify, not always attributed to mechanisms included in the model. We will address by either removing parts deemed superlative or speculative or adding explicit references to model processes.

These changes will be made to ‘4.2. Simulating grassland C, N and P pools by varying plant access to P sources’ in the discussion section.

‘On a more general topic, why simulations over such a long time period?’

The N14CP model is spun up from the onset of the Holocene to capture the length of time required for soil formation following deglaciation. This is not in an attempt to truly model this long time period but to form an initial condition for modern day simulations that takes in what we know about the site history and forcings. We prefer this method over spinning up a model over an undefined time period until it matches a SOC measurement, as is common practice with other similar models, as it avoids the assumption that soils are presently in steady state (which they are not), and the biasing of results from tuning to that initial stock. If after the spin up period used here, the model can simulate the magnitude of contemporary soil C, N and P pools, it’s a good indicator that the processes used by the model and its calibration of initial conditions (P_{Weath0} for example) is suitably reflective of our empirical data.

In addition, N14CP runs on a quarterly time-step and is therefore well-suited to simulating timescales from decades to centuries, which is beneficial considering the timescales of changes in soil pool conditions and nutrient stocks, and responses to long term changes in nutrient availability.

If this comment is more in reference to the timeseries presented in figure 4 then we believe simulating from 1800, and the onset of large scale N deposition across Europe, is the best starting point for investigating the consequences of N deposition on ecosystem C, N and P pools. Additionally, this timescale allows the effect of more recent nutrient treatments to be visually compared to background N deposition effects more clearly.

Thank you for asking this, we expect this to be question other readers may have and so it would be helpful if we were to add some justification into the methodology section. We first mention the timescale in ‘2.2.2. Net primary productivity’ so this would be an appropriate place for more information.

‘The effects of acidification on P availability regarding iron and aluminum complexes are more complex than referenced on lines 541-545 (see Barrow 2020). More information about how these sites, their mineralogy and pH might influence P availability would help interpret this idea.'

Thank you for raising this point and for providing a useful reference to help build upon it. We shall add more information about other soil factors that may influence P availability.

‘In conclusion: this model revision seems to have improved the N14CP model’s ability to respond to N and P limitations to plant growth, likely due to adding an organic P source, but the model doesn’t capture much of the soil pool dynamics so it could be summarized in a much shorter article.’

We hope that our previous comments have somewhat justified the detail provided in the article, though we acknowledge that we can reduce areas that may be deemed speculative (parts of the discussion) and that other sections may require additional information (model methods).